

Planning Sensitivities for Building Contingency Robustness and Graph Properties into Large Synthetic Grids

Adam B. Birchfield and Thomas J. Overbye

Department of Electrical and Computer Engineering, Texas A&M University

birchfieldllc@gmail.com, overbye@tamu.edu

Abstract

Interest in promoting innovation for large, high-voltage power grids has driven recent efforts to reproduce actual system properties in synthetic electric grids, which are fictitious datasets designed to be large, complex, realistic, and totally public. This paper presents new techniques based on system planning sensitivities, integrated into a synthesis methodology to mimic the constraints used in designing actual grids. This approach improves on previous work by explicitly quantifying each candidate transmission line's contribution to contingency robustness, balancing that with geographic and topological metrics. Example synthetic grids built with this method are compared to actual transmission grids, showing that the emulated careful design also achieves observed complex network properties. The results shed light on how the underlying graph structure of power grids reflects the engineering requirements of their design. Moreover, the datasets synthesized here provide researchers in many fields with public power system test cases that are detailed and realistic.

1. Introduction

The need for open data for reproducible science has been stressed by funding agencies and publishers in many fields, recognizing that allowing greater access to data that underlies a research activity helps enhance the robustness of results and stimulate further innovation [1]-[3]. For the field of power engineering, this need is especially pronounced because cyber and physical security concerns result in heavy restrictions on much electric grid data. In the United States, most detailed power grid models are considered Critical Energy Infrastructure Information (CEII), accessible only to regulators, utilities, and some researchers under non-disclosure agreements (NDAs) [4].

This need has recently led to growing interest in the development of more public datasets and test cases, as highlighted in a recent report of the National Academy

of Engineering [5]. Synthetic grids are fictitious datasets that do not contain CEII and are similar to actual datasets in size, complexity, and characteristics. Previous work in this area has produced grid topologies [6]-[8] with applications such as optimization [9], dynamics [10], and education [11]. The goal in building synthetic grids is to spur innovation by providing publishable test cases to research communities across multiple disciplines that are associated with the electric grid. In the power systems discipline, where most results are published using systems which are either simplistic or confidential, new synthetic grids enhance reproducibility of results and scientific rigor.

The focus of this paper is to present an improved methodology for constructing synthetic transmission networks that matches relevant graph properties by mimicking actual transmission system planning in balancing geographic and electric reliability constraints. The methodology in this paper builds on previous methods such as [7] by using a planning sensitivity technique that quantifies the impact of each candidate transmission line on the contingency robustness of the transmission system. With three large grids created, the paper uses observations made about graph properties of actual, confidential North American grids, compared directly with results for comparable synthetic grids. The results show that these synthetic grids, in pursuing the balance between geographic and reliability constraints, also demonstrate the observed complex network properties. Thus the resulting grids are not only well-suited as engineering testbeds, but match actual grids very well from a complex networks point of view. Hence these results provide insight into the dynamics associated with the structure of power grids.

2. Background on power grid complex network analysis

Much work in the area of synthetic grids has proposed methods that rely on complex network theory and graph generation techniques [6]-[8]. Studying the complex network properties of the electric grid predates the synthetic grid problem and has broader implications, particularly for vulnerability studies to random or

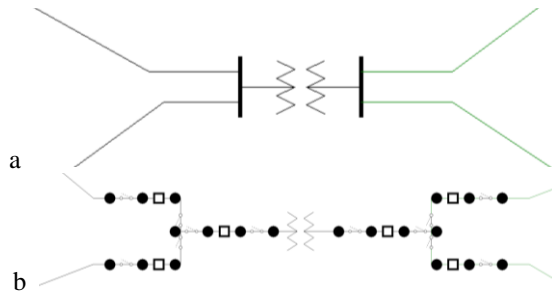


Figure 1. Effect of modeling detail on network topology metrics. This substation can be modeled with two buses (a) or with a full topology of 20 nodes (b). With all switches and breakers closed, these two models are electrically identical, but expanding to a full topology representation changes the average vertex degree from 3.0 to 2.1 by inserting many new degree-two vertices.

targeted attacks and the potential for cascading outages [12]-[15].

The literature lacks consensus on the complex network properties of large electric grids; this problem is exacerbated by the difficulty of accessing and sharing high-quality datasets for these mammoth machines. A survey from a few years ago showed that the reported average nodal degree of transmission systems ranged from 2.12 to 4.38, for example [16]. Some studies have touted the small-world network model as representing grid structure well [17], [6]; at least one has suggested a modified scale-free model [8]; others have rejected both [7], [13]. In addition to the influence of data non-availability, there is diversity among power grid models,

stemming from historical engineering design decisions, particular needs of various locations, and different levels of modeling detail. Many previous studies did not fully account for these considerations, which may contribute to the frequent discrepancies. So this paper supplements existing literature's findings with new analysis of three non-public power grid planning cases, for comparing to the synthetic datasets of this paper with about the same size and geographic footprint. The three systems observed, which comprise much of North America, are the eastern interconnect (EI, modeled with about 70,000 buses), the western interconnect (managed by the Western Electric Coordinating Council, or WECC, modeled with about 20,000 buses), and the Texas interconnect (managed by the Electric Reliability Council of Texas, or ERCOT, modeled with about 5000 buses).

A critical first decision is how the systems, real or synthetic, are viewed as a graph. As Figure 1 indicates, while the electrical circuit is a graph-based model with circuit nodes (buses) connected by branch elements (edges), there are multiple ways to model the same system which may be electrically equivalent but are not topologically identical. Another important consideration is whether generators are modeled at the transmission bus or behind their own step-up transformer, which would add many radial vertices with degree one. These subtle distinctions in modeling may dramatically affect graph metric properties, while the electric analysis remains minimally affected. To minimize these concerns and get to the core of the power system structure, this paper considers each substation as a combined vertex, with edges being actual transmission lines that connect two substations in a single section. For the synthetic cases, the substation identities are known;

Table 1. Algorithm steps for building synthetic grids.

Stage	Algorithm Steps
I	Substation Planning: use public input data on load and generation to set up synthetic substations
II	Transmission Planning: connect the substation set with dc-validated network of transmission lines 1) Initialize with randomized set of the correct number of lines 2) Iterate: a) Random removal from each sub-net b) Graph theory and geographic analysis c) Power flow N-1 contingency analysis and sensitivities d) Strategic addition to target objectives
III	Reactive Power Planning: iteratively transition from dc to ac power flow with associated reactive power devices

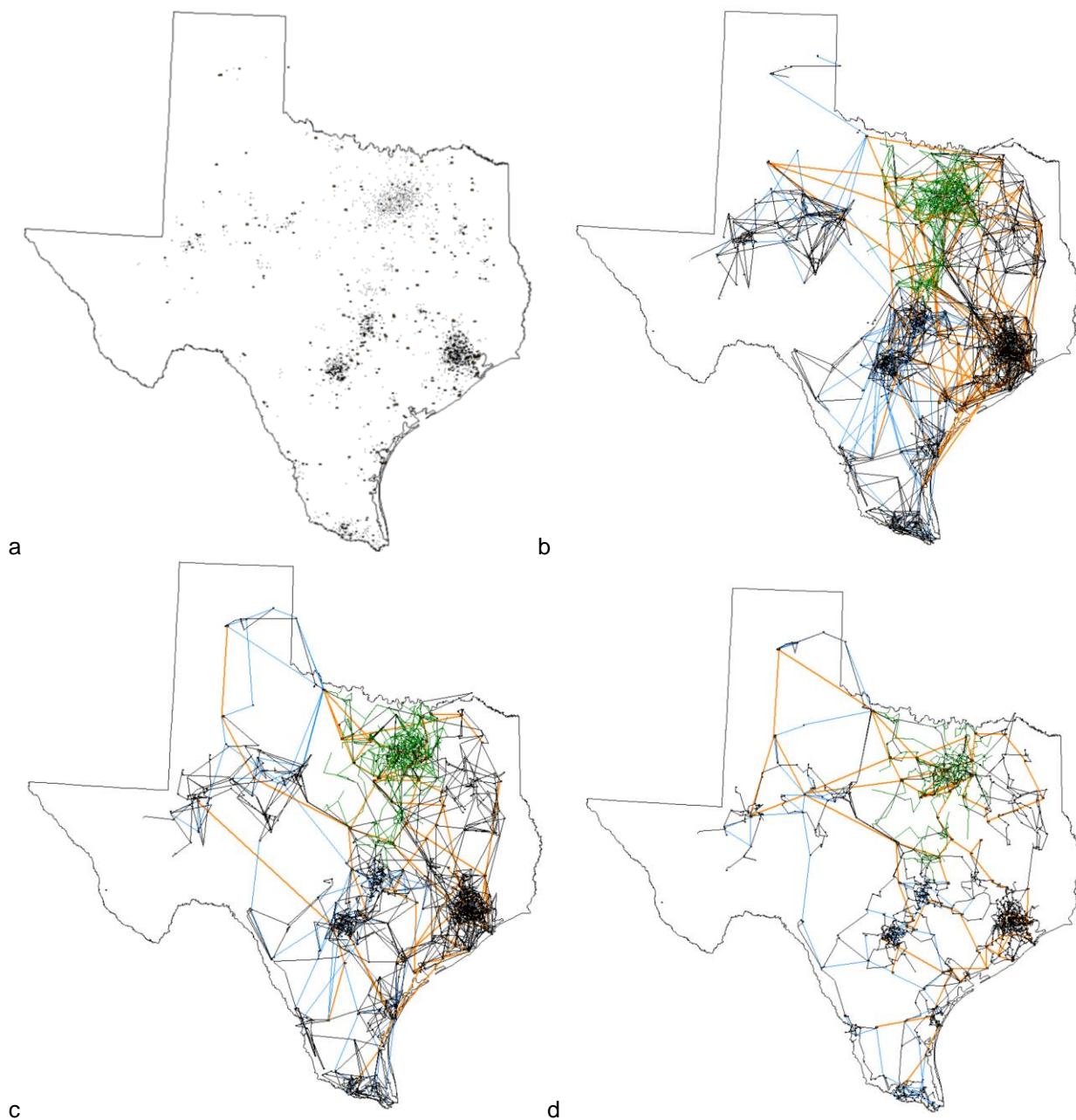


Figure 2. Stages of the transmission planning process in synthetic grid creation, shown on the 5000 bus case. a, The starting point is the geographic placement of substations. b, the grid is initialized with a random subset of $1.2n$ of the $21n$ candidate transmission lines. c, After 100 iterations of random removal followed by targeted addition, the grid begins to match more geographic and reliability constraints. d, After 10000 iterations, nearly all reliability and geographic constraints are met together. The line color represents the nominal voltage level.

for the actual data cases they sometimes need to be inferred.

3. Framework for building synthetic transmission grids

This paper's approach to building synthetic grids is mainly inspired by engineering goals of actual transmission planning, which are to balance the need for electric reliability against geographic constraints [18]. As in previous work, the substations of these synthetic cases are formed from seeds of public information relating to generating stations and population in a targeted geographic area [19], [20]. Like actual power system planning, the synthetic transmission grid design is preceded by load forecasting and planning out generators and substations. The substations are assigned nominal voltage levels with a clustering technique; then they are connected internally with transformers and low-impedance branches. Because this approach is geographic from the start, the approximate length of any potential transmission line between substations is known, and, combined with a known nominal voltage level, a realistic circuit impedance and rated power flow limit can be assigned. Throughout, statistics observed from actual power systems are used to set the values of electrical properties of the system. For synthesis purposes, each substation is considered to be a vertex in both the full, combined grid and one or more subnets, which comprise the interconnect and contain less than 1000 substations each at a particular voltage level in a geographic area. Though there are $n \cdot (n - 1)$ possible transmission lines connecting n substations at a voltage level, previous work has shown that nearly all practical transmission lines are in the tractable set of about $21n$ lines that comprise the geographic Delaunay triangulation and its second and third neighbors [7]. These are the candidate lines, and $1.15n$ to $1.25n$ of them must be chosen to form the synthetic network.

In contrast to previous work, this paper employs a new procedure for building the transmission network of a synthetic power grid, summarized in Table 1 and illustrated by the 5000-bus synthetic Texas case in Figure 2. The grid is initialized with an arbitrary selection of lines. At each iteration, remove one branch at random from each subnet, and then pick a candidate line that best contributes to the goals for the transmission system, and add it back to the network. Inspired by simulated annealing, these repetitive steps of random removal, smart addition, produce a network which balances both key objectives in actual transmission system planning: geographic feasibility and electric reliability.

The geographic feasibility goal considers, primarily, line length. Cost-effective planning prefers shorter lines, so the proposed method gives candidate lines a piecewise-linear penalty for length, gently encouraging shorter lines within the reasonable range for a nominal voltage level and sharply disallowing lines longer than are actually seen. The candidate line's length is calculated from the straight-line path between

substations, scaled by any geographic features it crosses. In addition to length, the synthesis process works to match the distribution of Delaunay neighbor count: there should be the right proportion of lines in the system that are first, second, and third neighbors along the Delaunay graph, matching actual grids [7]. To enforce this, uniform penalties are given to candidate lines in a category which is already overrepresented in the system.

4. Modeling reliability goals with transmission planning sensitivities

The electric reliability analysis for the proposed method focuses on the key planning criteria that electric service should not be interrupted, nor should lines be overloaded, by the outage of any single element (the N-1 criterion [21]). To achieve this, there is both a topological analysis and a power flow analysis. The topological analysis performs a depth-first graph search, returning the connected components, bridges, articulation points, and biconnected components. The graph should be fully connected and remain so with the loss of a single edge or vertex. With negative penalties, candidate transmission lines are strongly encouraged between two disconnected components and those between two substations that have only a single path between them. In later iterations, bridges are protected against removal to ensure connectivity.

The power flow part of the electric reliability analysis uses linearized (or "dc") power flow modeling, which ignores the flow of reactive power and assumes voltage magnitudes are equal to the nominal value [22]. This modeling mimics the analysis done in initial studies for actual transmission system planning. For each synthesis iteration, a full N-1 contingency set is run. Then for each subnet a critical contingency is chosen which causes the most branch overloads. To encourage system N-1 reliability, transmission lines are favored which contribute to easing these critical contingency overloads. For a few critical contingency overloads, the sensitivity of all $21n$ candidate lines to the power flowing in the overloaded line can be calculated quickly, as reviewed below [23].

The sensitivity metric used for ranking candidate lines is modified slightly from its original formulation [23], but it is based on the assumptions of the dc power flow. Using the \mathbf{B} matrix dc power flow formulation and differentiating it to a small change in system impedance:

$$\bar{P} = \mathbf{B} \cdot \bar{\theta} \quad (4)$$

$$d\bar{\theta} = -\mathbf{B}^{-1}(d\mathbf{B})\bar{\theta} = -\mathbf{B}^{-1}\bar{e}_i\bar{e}_i^T\bar{\theta}dB_i, \quad (5)$$

where the \mathbf{B} matrix is being modified on right-of-way i , where \bar{e}_i defines the right-of-way path by being zero at all buses except a 1 at the from bus and -1 at the to bus of the right-of-way. The \mathbf{B} bus matrix, which would reflect the system configuration in a given contingency, would be modified with only four values by the differential change in admittance dB_i . Now (5) gives the sensitivity of system angles to that admittance change. Denoting right-of-way k impedance as X_k and angle difference as ϕ_k , to get the change in power across it by the dc power flow approximations,

$$P_k = X_k \phi_k = X_k \bar{e}_k^T \bar{\theta} \quad (6)$$

$$d\phi_k = \bar{e}_k^T d\bar{\theta}. \quad (7)$$

Thus the sensitivity formulation becomes

$$\frac{\partial \phi_k}{\partial B_i} = -\phi_i \bar{e}_k^T \mathbf{B}^{-1} \bar{e}_i. \quad (8)$$

Since the \mathbf{B} matrix is symmetric, one can then rearrange the sensitivity to be

$$\frac{\partial \phi_k}{\partial B_i} = -\phi_i \bar{e}_i^T \mathbf{B}^{-1} \bar{e}_k. \quad (9)$$

So with one LU factorization of \mathbf{B} and backward substituting for \bar{e}_k , the sensitivity of the power through that right-of-way k can be determined with respect to adding admittance to any other right-of-way i by only three floating-point operations each, as follows:

$$\bar{s}_k = \mathbf{B}^{-1} \bar{e}_k \quad (10)$$

$$\frac{\partial \phi_k}{\partial B_i} = -\phi_i \bar{e}_i^T \bar{s}_k = -(\bar{\theta}[i_{\text{from}}] - \bar{\theta}[i_{\text{to}}]) \cdot (\bar{s}_k[i_{\text{from}}] - \bar{s}_k[i_{\text{to}}]) \quad (11)$$

Then the candidate lines are given a penalty negatively proportional to their sensitivity values, which encourages adding lines to mitigate these critical contingency overloads.

Figure 3 illustrates the use of the contingency sensitivity metrics during the synthesis process. This is zoomed in to a coastal area of the 5000-bus case partway through the grid synthesis. The black lines show the present state of the system. During the contingency analysis, it was found that during a contingency which outages one existing line (magenta) another existing line (orange) is overloaded. This contingency violation is targeted to be fixed, and the sensitivity vector (10) is calculated for the given contingency relative to the orange line (k). Then the impact of every candidate line (green and red in Figure 3) on this violation can be quantified using (11). On the figure, darker green lines do the most to reduce the overload and red lines make the overload worse. These sensitivities will be balanced against geographic and topological constraints to determine which line to add for the next iteration. The impact depends not only on the sensitivity but also the

Table 2. Summary of complex network properties real and synthetic.

Metric	Actual Systems			Synthetic Systems		
	EI	WECC	ERCOT	70K	20K	5000
n	36,187	9398	3827	34,999	11,765	2941
\bar{d}	2.61	2.58	2.61	2.74	2.99	3.12
\bar{c}	0.044	0.058	0.032	0.048	0.071	0.089
\bar{l}	29.2	18.9	14.2	36.7	22.0	13.7
\bar{b}	0.083	0.21	0.40	0.11	0.20	0.50

magnitude of the admittance added, which would be much more for a shorter line. Many of the longer lines, though the violation is highly sensitive to them, might not be added because of their length and other geographic constraints. This method improves on a two-stage approach used in previous work that focused first on the base case and then on additional lines needed for N-1 security [24].

5. Results on North American grids

Using this approach, three test grids were built with the analysis methodology of this paper, corresponding to the approximate geography and size of the actual grids analyzed, three portions of North America.

Geographic, single-line circuit diagrams of a resulting 70,000 bus synthetic system on the footprint of the U.S. portion of the EI can be seen in Figure 4. In addition, a 20,000 bus system, which is on the footprint of the U. S. portion of WECC, and a 5000 bus system, located on the ERCOT footprint, were constructed. In Figure 4, the transmission lines are shown, colored by voltage level. The geographic footprint uses states as areas and considers features including coastlines, mountain ranges, and urban centers. Arrows in the figure indicate the magnitude and direction of power flow along the lines.

Summary metrics are given in Table 2, for the three actual system datasets and the corresponding synthetic datasets. Each power flow dataset was pre-processed to identify the substations as single vertices, connected by transmission lines. n is the number of substation vertices. The average vertex degree \bar{d} can be calculated as $\bar{d} = 2m/n$, where m is the number of transmission line edges. The Watts-Strogatz clustering coefficient \bar{c} is calculated by averaging, for each vertex, the fraction of possible connections between neighbors that actually exist [17]. Vertices with degree 1 are ignored for the purpose of calculating \bar{c} , since there are no possible interconnections between pairs of neighbors. The average shortest path length \bar{l} is the average number of

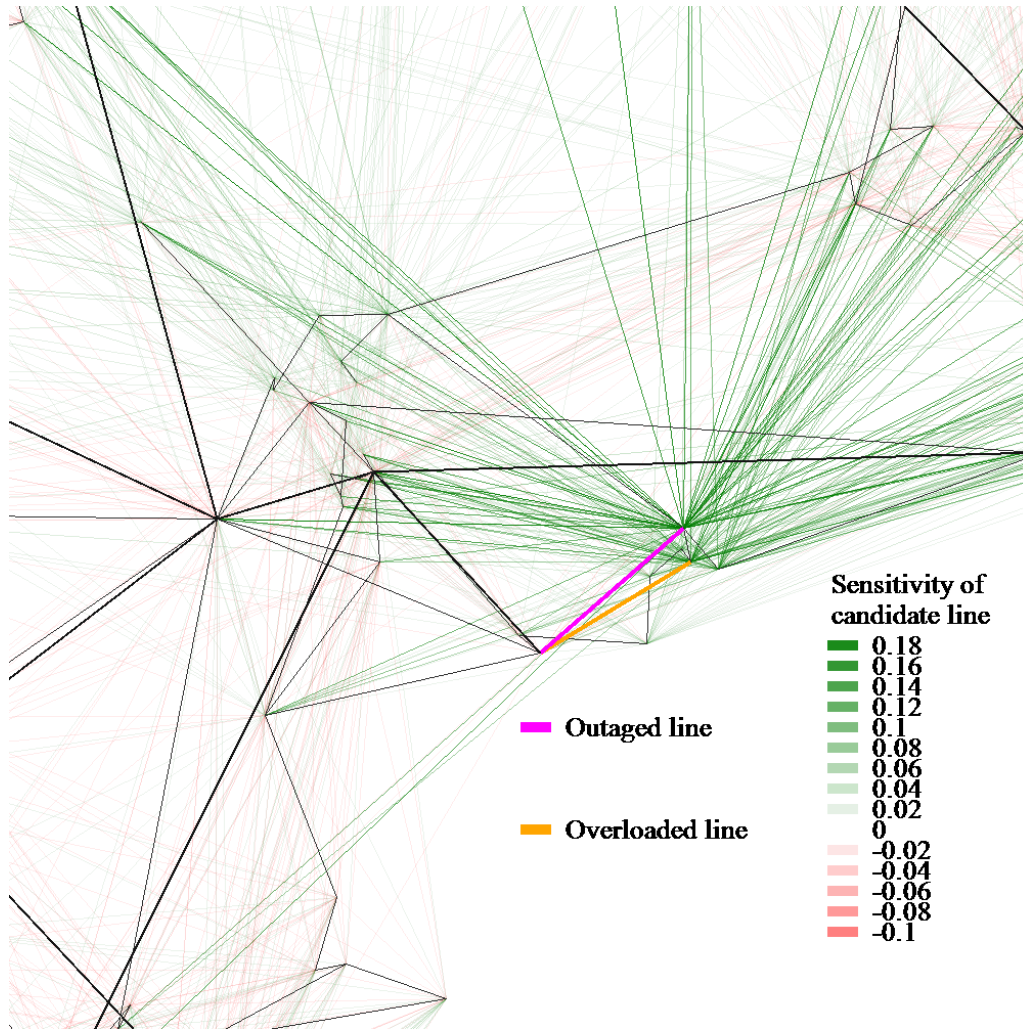


Figure 3. Reliability sensitivity visualization for a portion of the 5000 bus case, partway through the synthesis process. (The units of the sensitivity are per-unit power / per-unit admittance.)

hops between any two pairs of vertices. The average betweenness centrality quantifies what percentage of shortest paths go through the average vertex.

The systems' substation vertices correspond to about two buses each on average. The average degree \bar{d} is in effect a design parameter, since it relates the number of lines to the number of substations. But the actual systems show remarkable consistency, within 1% of an average degree of 2.6 for all three North American grids. This value fits comfortably in the range reported by literature and validates a design choice for the synthetic grids: how many transmission lines can be placed to meet the other objectives.

The degree distribution has been frequently discussed, and the data observed here supports the common notion that an exponential distribution fits the

data well, as shown in Figure 5, with the prominent exception that there are fewer vertices with degree one (radial substations) than an exact exponential distribution would predict. These results show degree two vertices to be the most common kind. This degree distribution confirms that power grids are not scale-free; a degree distribution for a scale-free network appears linear when both the horizontal and vertical axes are logarithmic [12].

For Watts and Strogatz, small world networks are characterized by a combination of average clustering coefficient and average shortest path length [17], and others have corroborated an observation of power grids that matches the Watts-Strogatz conditions [6]. From Table 2, the clustering coefficients are high, even relative to previously published numbers, with the

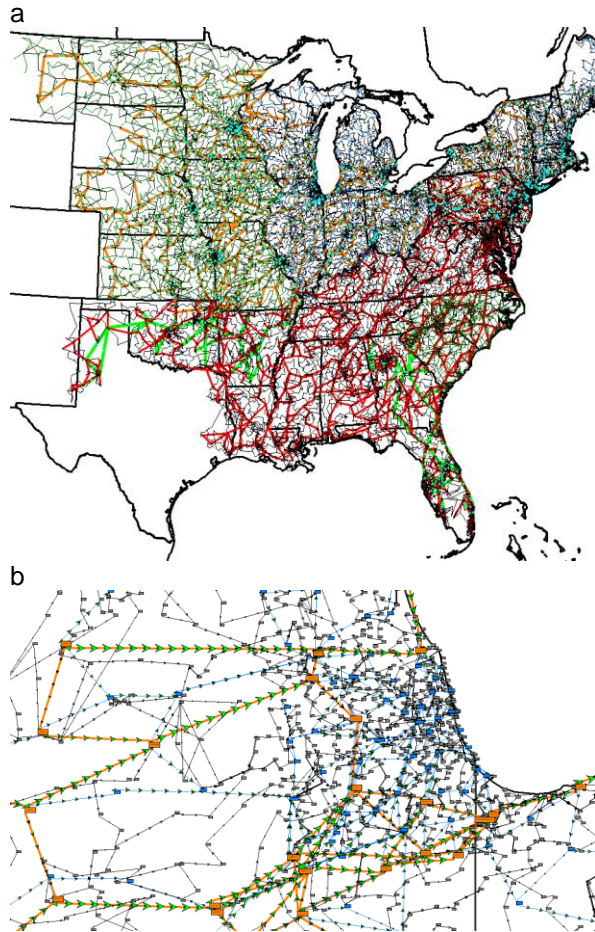


Figure 4. Single line diagram of the 70,000 bus synthetic power grid. a, The full interconnect overview. b, Zoomed-in view of the Chicago metropolitan area. This system is fictitious and does not represent any actual power grid.

average node having 3-6% of its neighbors interconnected. More detailed examination of this data reveals, however, that over 85% of nodes have a clustering coefficient of zero.

The average shortest path lengths are quite small, in the range of previous reports and small world conditions. However, what [17] emphasizes is that it is the scaling with respect to network size that is most interesting. This scaling is difficult to study when so few power grid networks actually exist at a variety of scales. The observations in Table 2 show that the scaling is decidedly sub-linear. While the EI has about 10 times as many substation vertices as ERCOT, its average shortest path length is little more than twice as large. This is a

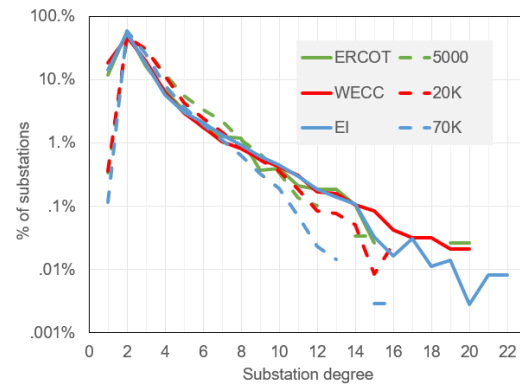


Figure 5. Degree distribution for real and synthetic grids showing the probability distribution function for the number of transmission line edges connecting to each substation vertex. Color indicates the size of the case; the solid lines are for actual grids and the dashed ones are for synthetic.

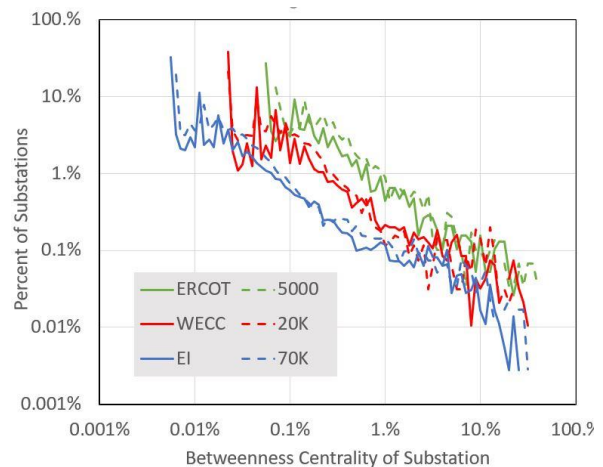


Figure 6. Betweenness centrality distribution for real and synthetic grids, showing the probability distribution function for the fraction of shortest paths routed through each substation vertex. Color indicates the size of the case; the solid lines are for actual grids and the dashed ones are for synthetic.

main reason the small world model is insufficient to describe the network structure of power grids.

Betweenness centrality is a metric associated with the topological routes of the shortest paths between pairs of vertices. For any vertex, the fraction of these n^2 paths that passes through it is that vertex's betweenness centrality. The vertices that score highly in this metric

are central topologically, placed between many pairs of nodes. This distribution is best shown on a log-log plot in Figure 6, with most substations having a very low betweenness centrality and a smaller number of key central substations. In general, the smaller grids have a more rightward-shifted distribution since a higher proportion of paths will pass through the typical vertex. In every case, the synthetic grid distribution lies almost directly on top of the actual case.

6. Conclusions

This paper presented a methodology based on power system planning sensitivities to balance geographic constraints against reliability criteria in constructing synthetic grid datasets. The sensitivity approach allows tractable analysis of each candidate line to quantify the benefit to reliability of each addition. Three example systems are built here and compared to similarly sized and located actual grids in North America, showing the similarity between the synthetic and actual grids.

7. References

- [1] IEEE and NSF. "Report on the first IEEE workshop on the future of research curation and research reproducibility (2016)." [Online]. Available: https://www.ieee.org/content/dam/ieee-org/ieee/web/org/ieee_reproducibility_workshop_report_final.pdf
- [2] National Institutes of Health (NIH). "NIH strategic plan for data science (2018)." [Online]. https://commonfund.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf
- [3] B. A. Nosek, et al. "Promoting an open research culture," in *Science*, vol. 348, no. 1, pp. 1422-1425, 2015.
- [4] Federal Energy Regulatory Commission (FERC). *FERC Order 630-A: Critical Energy Infrastructure Information (2003)*. [Online]. <https://www.ferc.gov/legal/maj-ord-reg/land-docs/ceii-rule.asp>
- [5] The National Academies of Sciences, Engineering, and Medicine. *Enhancing the resilience of the nation's electric system*. The National Academies Press, Washington, 2017.
- [6] Z. Wang, A. Scaglione, and R. J. Thomas. "Generating statistically correct random topologies for testing smart grid communication and control networks," in *IEEE Transactions on Smart Grid*, vol. 1, pp. 28-39, 2010.
- [7] A. B. Birchfield, T. Xu, K. M. Gegner, K. S. Shetye, and T. J. Overbye. "Grid structural characteristics as validation criteria for synthetic networks," in *IEEE Transactions on Power Systems*, vol. 32, pp. 3258-3265, 2017.
- [8] S. Soltan, A. Loh, and G. A. Zussman. "Learning-based method for generating synthetic power grids," in *IEEE Systems Journal*, vol. 13, pp. 625-634, 2018.
- [9] S. Misra, L. Roald, and Y. Ng. "Learning for convex optimization: Identifying optimal active constraint sets," pre-print 2019. [Online]. <https://arxiv.org/pdf/1802.09639.pdf>
- [10] T. Xu, A. B. Birchfield, and T. J. Overbye. "Modeling, tuning, and validating system dynamics in synthetic electric grids," in *IEEE Transactions on Power Systems*, vol. 33, pp. 6501-6509, 2018.
- [11] A. B. Birchfield, T. J. Overbye, and K. R. Davis. "Educational applications of large synthetic power grids," in *IEEE Transactions on Power Systems*, vol. 34, pp. 765-772, 2018.
- [12] R. Albert, I. Albert, and G. L. Nakarado. "Structural vulnerability of the North American power grid," in *Phys. Rev. E*, vol. 69, pp. 025103, 2004.
- [13] E. Cotilla-Sanchez, P. D. Hines, C. Barrows, and S. Blumsack. "Comparing the topological and electrical structure of the North American electric power infrastructure," in *IEEE Systems Journal*, vol. 6, pp. 616-626, 2012.
- [14] Y. Yang, T. Nishikawa, and A. E. Motter. "Small vulnerable sets determine large network cascades in power grids," in *Science*, vol. 358, pp. 886, 2017.
- [15] B. Shafer, D. Witthaut, M. Timme, and V. Latora. "Dynamically induced cascading failures in power grids," in *Nature Communications* vol. 9, pp. 1975, 2018.
- [16] G. A. Pagani and M. Aiello. "The power grid as a complex network: A survey," in *Physica A*, vol. 392, pp. 2688-2700, 2013.
- [17] D. J. Watts and S. H. Strogatz. "Collective dynamics of 'small-world' networks," in *Nature* vol. 393, pp. 440-442, 1998.
- [18] H. Seifi and M. S. Sepasian. *Electric power system planning*. Springer, 2011.
- [19] United States Energy Information Administration. *2015 Form 860: Annual electric generator report*. [Online]. <http://www.eia.gov/electricity/data/eia860/>
- [20] United States Census Bureau. *2015 Census U. S. Gazetteer Files*. [Online]. <http://www.census.gov/geo/maps-data/data/gazetteer.html>
- [21] North American Electric Reliability Corporation (NERC). *Standard TPL-001-2 Transmission System Planning Performance Requirements*.

- [Online]. <https://www.nerc.com/files/TPL-001-2.pdf>
- [22] B. Stott, J. Jardim, and O. Alsac. "Dc power flow revisited," in *IEEE Transactions on Power Systems*, vol. 24, pp. 1290-1300, 2009.
- [23] R. J. Bennon, J. A. Juves, and A. P. Meliopoulos, "Use of sensitivity analysis in automated transmission planning, in *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-101, pp. 53-59, 1982.
- [24] A. B. Birchfield, H. Li, and T. J. Overbye, "Security considerations in transmission planning for creating large synthetic power grids," *Clemson University Power Systems Conference*, Charleston, South Carolina, Sept. 2018.
- [25] S. H. Strogatz. "Exploring complex networks," in *Nature*, vol. 410, pp. 268-276, 2001.